Supplementary Appendix

Proposition 5.

(i) *Suppose that $\tau \in \mathcal{C}$ is conditionally independent. If $\mu = marg_{\Theta} p_\tau$ is not degenerate, then τ is minimal iff it is not perfectly informative. If μ is degenerate, then τ is minimal.*

(ii) *A public $\tau \in \mathcal{C}$ is minimal iff $\operatorname{supp} \tau$ is a singleton.*

Proof. Part (i). Suppose that τ is conditionally independent and μ is non-degenerate. First, if τ is perfectly informative, then it can be written

$$\tau = \sum_{\theta} \mu(\theta) \tau_\theta,$$

where τ_θ is a distribution that gives probability 1 to belief hierarchies representing common knowledge that θ has realized. Given $\mu(\theta) \in (0,1)$, τ is therefore a convex combination of belief-hierarchy distributions, hence it is not minimal.

Second, we show that if τ is non-minimal, then it must be perfectly informative. Let τ be non-minimal. By Proposition 3, there exist $\alpha \in (0,1)$ and $\tau' \neq \tau''$ such that $\operatorname{supp} \tau' \cap \operatorname{supp} \tau'' = \emptyset$ and $\tau = \alpha\tau' + (1-\alpha)\tau''$.

Now, take $t' \in \operatorname{supp} \tau'$, $t'' \in \operatorname{supp} \tau''$ and note

$$p_\tau(t'_i, t''_{-i}|\theta) = \alpha p_{\tau'}(t'_i, t''_{-i}|\theta) + (1-\alpha) p_{\tau''}(t'_i, t''_{-i}|\theta) = 0 \tag{1}$$

for all θ and i. If τ were conditionally independent,

$$\begin{aligned} & p_\tau(t'_i, t''_{-i}|\theta) \\ = \; & p_\tau(t'_i|\theta) \textstyle\prod_{j\neq i} p_\tau(t''_j|\theta) \\ = \; & \left(\alpha p_{\tau'}(t'_i|\theta) + (1-\alpha) p_{\tau''}(t'_i|\theta)\right) \textstyle\prod_{j\neq i} \left(\alpha p_{\tau'}(t''_j|\theta) + (1-\alpha) p_{\tau''}(t''_j|\theta)\right), \end{aligned}$$

which is strictly positive for some θ when τ is not perfectly informative, and thus contradicts (1). This implies that a non-minimal conditionally independent τ must be perfectly informative.

Part (ii). If τ is public, then every $\{t\}$ such that $t \in \operatorname{supp} \tau$ is a consistent distribution. Therefore, if $\operatorname{supp} \tau$ is a singleton, then it is clearly minimal. But if $\operatorname{supp} \tau$ is not a singleton, then τ is a convex combination of multiple consistent distributions, in which case τ is not minimal. □

Next we show that $\mathcal{C}^{\text{M}}$ is small in a measure-theoretic sense relative to $\mathcal{C}$. Since there is no analog of the Lebesgue measure in infinite dimensional spaces, we use

the notion of finite shyness proposed by Anderson and Zame (2001), which captures the idea of Lebesgue measure 0.

Definition 3. A measurable subset A of a convex subset C of a vector space $\mathcal{S}$ is finitely shy if there exists a finite dimensional vector space $V \subseteq \mathcal{S}$ for which $\lambda_V(C+s) > 0$ for some $s \in \mathcal{S}$, and $\lambda_V(A+s) = 0$ for all $s \in \mathcal{S}$, where λ_V is the Lebesgue measure defined on V.

Proposition 6. *$\mathcal{C}^M$ is finitely shy in $\mathcal{C}$.*

Proof. By Lemma 1, $\mathcal{C}$ is a convex subset of the vector space $\mathcal{S}$ of all signed measures on T. Choose any distinct $e, e' \in \mathcal{C}^{\text{M}}$ and let $V = \{\alpha(e-e') : \alpha \in \mathbb{R}\} \subseteq \mathcal{S}$. By construction, V is a one-dimensional subspace of $\mathcal{S}$. Let $\lambda_V \in \Delta V$ represent the Lebesgue measure on V. Notice that $\alpha(e-e') = \alpha e + (1-\alpha)e' - s$ for $s := e'$ and that $(\mathcal{C}-s) \cap V = \{\alpha(e-e') : \alpha \in [0,1]\}$ by convexity of $\mathcal{C}$. Hence, $\lambda_V(\mathcal{C}-s) > 0$. However, since $\mathcal{C}^M$ is the set of extreme points of $\mathcal{C}$, for every $s \in \mathcal{S}$, $(\mathcal{C}^M - s) \cap V$ contains at most two points. This gives $\lambda_V(\mathcal{C}^M - s) = 0$, since points have Lebesgue measure zero in V. □

Proposition 7. *Let Σ be* BNE *and suppose that the selection criterion is* max. *For any minimal belief-hierarchy distribution $e \in \mathcal{C}^M$, there exist v and G for which $\lambda^* = \delta_e$ is the essentially unique optimal solution.*[1]

Proof. Fix $e \in \mathcal{C}^M$, $\varepsilon > 0$ and let $\mu_0 = p_e$. Denote by $G_\varepsilon(e) = (N, \{A_i, u_i\})$ the (base) game defined in Chen et al. (2010)'s Lemma 1 where $A_i \supseteq \operatorname{supp} e_i$. In this game, agent i's actions include belief hierarchies from e_i. The (base) game $G_\varepsilon(e)$ is so conceived that, in the Bayesian game $(G_\varepsilon(e), e)$, every agent has a strict incentive to truthfully report his true belief hierarchy, but for any τ that is suitably distant from e, some i in the Bayesian game $(G_\varepsilon(e), \tau)$ has a strict incentive *not* to report any $t_i \in \operatorname{supp} e_i$. For any i and $t_i, t_i' \in T_i$, let

$$d_i(t_i, t_i') := \sup_{k \geq 1} d^k(t_i, t_i')$$

where d^k is the standard metric (over k-order beliefs) that metrizes the topology of weak convergence. Let R_i and R be the ICR actions and profiles. Lemma 1 and

[1] The proof establishes the stronger claim with $\Sigma :=$ ICR by using a result from Chen et al. (2010). A minimal distribution e is the essentially unique optimal solution if for all $\varepsilon > 0$, there is a game G_ε such that all $\tau \in \mathcal{C}$ with $d(e, \tau) > \varepsilon$ are strictly suboptimal (the metric is defined in (2)). By choosing a constant game or a constant designer's utility, it is easy to make all minimal distributions optimal, since the designer is indifferent among them. Uniqueness makes the result much stronger.

Proposition 3 from Chen et al. (2010) imply that for every i and $t_i \in \operatorname{supp} e_i$,

$$t_i \in \operatorname*{argmax}_{a_i \in A_i} \sum_{\theta} \sum_{t_{-i} \in \operatorname{supp} e_{-i}} u_i(a_i, t_{-i}, \theta) \beta_i^*(\theta, t_{-i} | t_i)$$

and for every $\tau \in \mathcal{C}$ such that

$$d(e, \tau) := \max_i d_i^H(\operatorname{supp} e_i, \operatorname{supp} \tau_i) \geq \varepsilon, \tag{2}$$

where d_i^H is the standard Hausdorff metric, there exist i and $t_i' \in \operatorname{supp} \tau_i$ such that $\operatorname{supp} e_i \cap R_i(t_i') = \emptyset$. To see why, note that since e is minimal, there can be no sequence $(\tau_n) \subseteq \mathcal{C}$ such that $d(e, \tau_n) \geq \varepsilon$ for all n, while

$$\max_i \max_{t_i \in \operatorname{supp} \tau_{n,i}} \min_{t_i' \in \operatorname{supp} e_i} d_i(t_i, t_i') \to 0. \tag{3}$$

That is, for all τ such that $d(e, \tau) \geq \varepsilon$, there is $\delta > 0$ such that

$$\max_i \max_{t_i \in \operatorname{supp} \tau_{n,i}} \min_{t_i' \in \operatorname{supp} e_i} d_i(t_i, t_i') \geq \delta. \tag{4}$$

Put differently, there exist i and $t_i' \in \operatorname{supp} \tau_i$ such that $d_i(t_i, t_i') \geq \delta > 0$ for all $t_i \in \operatorname{supp} e_i$. From (the proof of) Proposition 3 in Chen et al. (2010), we conclude that $\operatorname{supp} e_i \cap R^k(t_i') = \emptyset$ for some k. Given that $R_i(t_i') = \cap_{k=1}^{\infty} R_i^k(t_i')$, we have $\operatorname{supp} e_i \cap R_i(t_i') = \emptyset$. Now, define the designer's utility as $v(a) := \mathbb{1}(a \in \operatorname{supp} e)$ for all $a \in A$. Then, the designer's expected payoff is

$$w(\tau) = \begin{cases} 1 & \text{if } \tau = e, \\ x & \text{if } d(e, \tau) < \varepsilon, \\ y & \text{if } d(e, \tau) \geq \varepsilon \end{cases}$$

where $x \leq 1$ and $y < 1$. When $d(e, \tau) < \varepsilon$, it is not excluded that $x = 1$, because all of $\operatorname{supp} \tau$, by virtue of being close to some hierarchy in $\operatorname{supp} e$, might report only in $\operatorname{supp} e$. However, whenever $d(e, \tau) \geq \varepsilon$, there is a hierarchy profile t occurring with positive probability that reports outside $\operatorname{supp} e$. Thus, the designer maximizes her expected payoff by setting $\lambda^* = \delta_e$, which is Bayes plausible since $\mu_0 = p_e$. □

References

Anderson, Robert M. and William Zame. 2001. "Genericity with Infinitely Many Parameters." *The B.E. Journal of Theoretical Economics* 1 (1):1–64.

Chen, Yi-Chun, Alfredo Di Tillio, Eduardo Faingold, and Siyang Xiong. 2010. "Uniform topologies on types." *Theoretical Economics* 5 (3).